

# On the First-Order Complexity of Induced Subgraph Isomorphism

Oleg Verbitsky\*

Maksim Zhukovskii†

## Abstract

Given a graph  $F$ , let  $\mathcal{I}(F)$  be the class of graphs containing  $F$  as an induced subgraph. Let  $W[F]$  denote the minimum  $k$  such that  $\mathcal{I}(F)$  is definable in  $k$ -variable first-order logic. The recognition problem of  $\mathcal{I}(F)$ , known as Induced Subgraph Isomorphism (for the pattern graph  $F$ ), is solvable in time  $O(n^{W[F]})$ . Motivated by this fact, we are interested in determining or estimating the value of  $W[F]$ . Using Olariu's characterization of paw-free graphs, we show that  $\mathcal{I}(K_3 + e)$  is definable by a first-order sentence of quantifier depth 3, where  $K_3 + e$  denotes the paw graph. This provides an example of a graph  $F$  with  $W[F]$  strictly less than the number of vertices in  $F$ . On the other hand, we prove that  $W[F] = 4$  for all  $F$  on 4 vertices except the paw graph and its complement. If  $F$  is a graph on  $\ell$  vertices, we prove a general lower bound  $W[F] > (1/2 - o(1))\ell$ , where the function in the little- $o$  notation approaches 0 as  $\ell$  increases. This bound holds true even for a related parameter  $W^*[F] \leq W[F]$ , which is defined as the minimum  $k$  such that  $\mathcal{I}(F)$  is definable in the infinitary logic  $L_{\infty\omega}^k$ . We show that  $W^*[F]$  can be strictly less than  $W[F]$ . Specifically,  $W^*[P_4] = 3$  for  $P_4$  being the path graph on 4 vertices.

## 1 Introduction

For a given graph  $F$ , let  $\mathcal{I}(F)$  denote the class of all graphs containing a copy of  $F$  as an induced subgraph. We are interested in the descriptive complexity of  $\mathcal{I}(F)$ , for a fixed pattern graph  $F$ , in first-order logic whose vocabulary consists of the adjacency and the equality relations ( $\sim$  and  $=$  respectively). Let  $D[F]$  denote the minimum quantifier depth of a sentence in this logic that defines  $\mathcal{I}(F)$ . Furthermore, let  $W[F]$  denote the minimum variable width of a sentence defining  $\mathcal{I}(F)$ . Note that

$$W[F] \leq D[F] \leq \ell,$$

---

\*Institut für Informatik, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin. Supported by DFG grant VE 652/1–2. On leave from the IAPMM, Lviv, Ukraine.

†Moscow Institute of Physics and Technology, Department of Discrete Mathematics, Moscow. Supported by grants No. 15-01-03530 and 16-31-60052 of Russian Foundation for Basic Research.

where here and throughout the paper  $\ell$  denotes the number of vertices in  $F$ . It may come as some surprise that the parameter  $D[F]$  can be strictly less than  $\ell$ . To see an example, let  $F = K_3 + e$  be the paw graph  $\text{paw}$ . The following sentence uses three variables  $x_1, x_2, x_3$  and has quantifier depth 3:

$$\begin{aligned} \exists x_1 (\exists x_2 \exists x_3 (x_1 \sim x_2 \wedge x_1 \sim x_3 \wedge x_2 \sim x_3) \wedge \\ \exists x_2 (x_1 \not\sim x_2 \wedge \exists x_3 (x_1 \sim x_3 \wedge x_3 \sim x_2) \wedge \exists x_3 (x_3 \sim x_1 \wedge x_3 \not\sim x_2))) ). \end{aligned}$$

It says that a graph contains a vertex  $v$  that belongs to a triangle and can be accompanied with a vertex  $u$  at distance 2 from  $v$  such that there is a vertex  $w$  adjacent to  $v$  but non-adjacent to  $u$ . Obviously, this sentence is true on the paw graph and on every graph containing an induced paw subgraph. Olariu's characterization of paw-free graphs [20] implies that the sentence is false on every graph without an induced paw subgraph; see Section 4.1 for details.

The decision problem for  $\mathcal{I}(F)$  is known as INDUCED SUBGRAPH ISOMORPHISM (for the pattern graph  $F$ ). We denote it by  $\text{ISI}(F)$ . Our interest in the parameters  $D[F]$  and  $W[F]$  is motivated by the fact that  $\text{ISI}(F)$  is solvable in time  $O(n^{W[F]})$ ; see [18, Prop. 6.6]. Before stating our results on  $D[F]$  and  $W[F]$  we very briefly overview the known algorithmic results in this area.

**Computational complexity of Induced Subgraph Isomorphism.** Obviously,  $\text{ISI}(F)$  is solvable in time  $O(n^\ell)$  on  $n$ -vertex input graphs by exhaustive search. We use the standard notation  $K_\ell$  for complete graphs,  $P_\ell$  for paths, and  $C_\ell$  for cycles on  $\ell$  vertices. Itai and Rodeh [13] observed that  $\text{ISI}(K_3)$  is solvable in time  $O(n^\omega)$ , where  $\omega < 2.373$  is the exponent of fast square matrix multiplication [11]. Nešetřil and Poljak [19] showed, by a reduction of  $\text{ISI}(F)$  to  $\text{ISI}(K_3)$ , that  $\text{ISI}(F)$  is solvable in time  $O(n^{(\omega/3)\ell+2})$ . On the other hand,  $\text{ISI}(K_\ell)$  is unsolvable in time  $n^{o(\ell)}$  unless the Exponential Time Hypothesis fails [3]. Floderus et al. [10] collected evidence in favour of the conjecture that  $\text{ISI}(F)$  for  $F$  with  $\ell$  vertices cannot be solved faster than  $\text{ISI}(K_{c\ell})$ , where  $c < 1$  is a constant independent on  $F$ . Along with the Exponential Time Hypothesis, this would imply that the time complexity of  $\text{ISI}(F)$  is  $n^{\Theta(\ell)}$ . As an example of a particular result of [10] in this direction,  $\text{ISI}(P_{2a-1})$  is not easier than  $\text{ISI}(K_a)$ , and the same holds true for  $\text{ISI}(C_{2a})$ ; see also the earlier work [4].

The induced subgraph problem has been intensively studied for particular pattern graphs  $F$  with small number of vertices. Let  $\overline{F}$  denote the complement graph of  $F$  and note that at least one of the graphs  $F$  and  $\overline{F}$  is connected. Since  $\text{ISI}(F)$  and  $\text{ISI}(\overline{F})$  have the same time complexity, we can restrict our attention to connected pattern graphs. There are six such graphs on four vertices, namely  $K_4$ ,  $P_4$ ,  $C_4$ ,  $K_3 + e$ , the claw graph  $K_{1,3}$  ( $\text{Y}$ ), and the diamond graph  $K_4 \setminus e$  ( $\text{diamond}$ ). Corneil et al. [6] designed an  $O(n + m)$  time algorithm for  $\text{ISI}(P_4)$ , where  $m$  denotes the number of edges in an input graph. As noted in [16], the Olariu characterization of paw-free graphs reduces  $\text{ISI}(K_3 + e)$  to  $\text{ISI}(K_3)$ , showing that the former problem is also solvable in time  $O(n^\omega)$ . The same time bound is obtained by Vassilevska

Williams et al. [26] for the diamond graph  $K_4 \setminus e$  and, using randomization, for the other pattern graphs on 4 vertices except  $K_4$ . The best known time bound for  $\text{ISI}(K_4)$  is given by the methods of [8] and is currently  $O(n^{3.257})$  [26].

**Our results on the descriptive complexity of Induced Subgraph Isomorphism.** In Section 3 we prove a general lower bound  $W[F] > (1/2 - o(1))\ell$ , where the function in the little-o notation approaches 0 as  $\ell$ , the number of vertices in the pattern graph  $F$ , increases. Whether or not it can be improved remains an intriguing open question. Note that this bound leaves a hypothetical possibility that the time bound  $O(n^{W[F]})$  for Induced Graph Isomorphism can be better than the Nešetřil-Poljak bound  $O(n^{(\omega/3)\ell+2})$  for infinitely many pattern graphs  $F$ .

Our approach uses a connection to the  $k$ -extension property of graphs, that is well known in finite model theory; see, e.g., [24]. We define the *extension index* of  $F$ , denoted by  $E[F]$ , as the minimum  $k$  such that the  $k$ -extension property forces the existence of an induced copy of  $F$ . It is easy to show that  $W[F] \geq E[F]$ . Our results about the parameter  $E[F]$  may be interesting on its own. In particular, we show that  $E[F] \geq \chi(F)$ , where  $\chi(F)$  denotes the chromatic number of  $F$ .

In Section 4, we determine the values of  $D[F]$  and  $W[F]$  for all pattern graphs with at most 4 vertices. We prove that  $W[F] = 4$  for all  $F$  on 4 vertices except the paw graph and its complement.

Our lower bounds for  $W[F]$  hold true also for a related parameter  $W^*[F] \leq W[F]$ , which is defined as the minimum  $k$  such that  $\mathcal{I}(F)$  is definable in the infinitary logic  $L_{\infty\omega}^k$ . The only exception is  $F = P_4$ , the path on 4 vertices. In this case, we prove that  $W^*[P_4] = 3$  while  $W[P_4] = 4$ . This shows that  $W^*[F]$  can be strictly less than  $W[F]$ , that is, the infinitary logic can be more succinct when defining  $\mathcal{I}(F)$ .

In Section 5, we address a relaxation version of the parameter  $W[F]$ . Consider a simple example. Let  $D_v[F]$  be the minimum quantifier depth of a sentence  $\Phi$  defining  $\mathcal{I}(F)$  over *sufficiently large connected* graphs. That is, it is required that there is a number  $s$  such that  $\Phi$  correctly detects whether or not a graph  $G$  belongs to  $\mathcal{I}(F)$  only if  $G$  is connected and has at least  $s$  vertices. Whereas  $D_v[F] \leq D[F]$ , it is clear that  $\mathcal{I}(F)$  for a connected pattern graph  $F$  is still recognizable in time  $O(n^{D_v[F]})$ . Let  $F = P_3$ . As easily seen,  $P_3$ -free graphs are exactly disjoint unions of cliques. Therefore, connected  $P_3$ -free graphs are exactly the complete graphs, which readily implies that  $D_v[P_3] \leq 2$ , whereas  $D_v[P_3] = 3$ . As a further example, we remark that the existence of a *not necessarily induced* subgraph  $P_4$  can be defined over sufficiently large connected graphs with just 2 variables; see Appendix A.

We can go further and define  $W_{tw}[F]$  to be the minimum variable width of a sentence defining  $\mathcal{I}(F)$  over connected graphs  $G$  of *sufficiently large treewidth*  $tw(G)$ . As a consequence of Courcelle's theorem [7],  $\mathcal{I}(F)$  for a connected pattern graph  $F$  is recognizable in time  $O(n^{W_{tw}[F]})$ ; cf. the discussion in [25].

The above discussion motivates the problem of proving lower bounds for the parameter  $W_\kappa[F]$  which we define as the minimum variable width of a sentence defining  $\mathcal{I}(F)$  over graphs  $G$  of *sufficiently large connectedness*  $\kappa(G)$ . Note that  $W_\kappa[F] \leq W_{tw}[F] \leq W_v[F] \leq W[F]$ . We prove that  $W_\kappa[F] = W[F]$  for a large class

of pattern graphs  $F$ . We also prove a general lower bound  $W_\kappa[F] > (1/3 - o(1))\ell$  for all  $F$  on  $\ell$  vertices.

**Comparison to (not necessarily induced) Subgraph Isomorphism.** The SUBGRAPH ISOMORPHISM problem is very different from its induced version. For infinitely many pattern graphs  $F$ , SUBGRAPH ISOMORPHISM can be solved in time  $O(n^C)$  for a constant  $C$ . This follows from a result by Alon, Yuster and Zwick [1] who showed that the problem is solvable in time  $2^{O(\ell)} \cdot n^{tw(F)+1} \log n$ .

Let  $W(F)$  and  $D(F)$  be the analogs of  $W[F]$  and  $D[F]$  for the not-necessarily-induced case. Note that  $W(F) = D(F) = \ell$  as a consequence of the trivial observation that  $K_\ell$  contains  $F$  as a subgraph while  $K_{\ell-1}$  does not. Nevertheless,  $D_v[F]$  can be strictly smaller than  $\ell$  for some connected pattern graphs. Moreover,  $D_{tw}[F]$  can sometimes be arbitrarily small comparing to  $\ell$ . This is the subject of our preceding paper [25].

**Logic with numeric predicates.** In the present paper, we consider the most laconic first-order language of graphs whose vocabulary has only the adjacency and the equality relations. If we assume that the vertex set of a graph is  $\{1, 2, \dots, n\}$  and additionally allow arbitrary numerical relations, this richer logic captures the non-uniform  $\text{AC}^0$ ; see [12, 18]. Let  $W_{\text{Arb}}(F)$  denote the analog of the parameter  $W(F)$  (the not-necessarily-induced case) for this logic, and  $W_{\text{Arb}}[F]$  denote the analog of the parameter  $W[F]$  (the induced case). The known relations to circuit complexity [12, 23] imply that the (not necessarily induced) SUBGRAPH ISOMORPHISM is solvable by bounded-depth unbounded-fan-in circuits of size  $n^{W_{\text{Arb}}(F)+o(1)}$ , and the similar is true also for INDUCED SUBGRAPH ISOMORPHISM. Whereas the parameter  $W_{\text{Arb}}(F)$  is studied in this context by Li, Razborov, and Rossman [17], the parameter  $W_{\text{Arb}}[F]$  remains unexplored.

## 2 Preliminaries

A *graph property* is a class of graphs  $\mathcal{C}$  closed under isomorphism, that is, for isomorphic graphs  $G$  and  $H$ ,  $G \in \mathcal{C}$  iff  $H \in \mathcal{C}$ . We consider first-order sentences about graphs in the language containing the adjacency and the equality relations. A sentence  $\Phi$  *defines* a graph property  $\mathcal{C}$  if  $G \in \mathcal{C}$  exactly when  $G \models \Phi$ , i.e.,  $\Phi$  is true on  $G$ . A graph property  $\mathcal{C}$  is *first-order definable* if there is a first-order sentence defining  $\mathcal{C}$ .

Let  $\mathcal{C}$  be a first-order graph property. The *logical depth* of  $\mathcal{C}$ , denoted by  $D(\mathcal{C})$ , is the minimum quantifier depth (rank) of a sentence defining  $\mathcal{C}$ . The *logical width* of  $\mathcal{C}$ , denoted by  $W(\mathcal{C})$ , is the minimum *variable width* of a sentence defining  $\mathcal{C}$ , i.e., the number of first-order variables occurring in the sentence where different occurrences of the same variable do not count.

Given two non-isomorphic graphs  $G$  and  $H$ , let  $D(G, H)$  (resp.  $W(G, H)$ ) denote the minimum quantifier depth (resp. variable width) of a sentence that *distinguishes*

$G$  and  $H$ , i.e., is true on one of the graphs and false on the other.

**Lemma 2.1.**  $D(\mathcal{C}) = \max \{ D(G, H) : G \in \mathcal{C}, H \notin \mathcal{C} \}.$

*Proof.* In one direction, note that whenever  $G \in \mathcal{C}$  and  $H \notin \mathcal{C}$ , we have  $D(G, H) \leq D(\mathcal{C})$  because any sentence defining  $\mathcal{C}$  distinguishes  $G$  and  $H$ . For the other direction, suppose that every such  $G$  and  $H$  are distinguished by a sentence  $\Phi_{G,H}$  of quantifier depth at most  $d$ . Specifically, suppose that  $\Phi_{G,H}$  is true on  $G$  and false on  $H$ . We have to show that  $\mathcal{C}$  can be defined by a sentence of quantifier depth at most  $d$ . For a graph  $G \in \mathcal{C}$ , consider the sentence  $\Phi_G \stackrel{\text{def}}{=} \bigwedge_H \Phi_{G,H}$ , where the conjunction is over all  $H \notin \mathcal{C}$ . This sentence distinguishes  $G$  from all  $H \notin \mathcal{C}$  and has quantifier depth at most  $d$ . The only problem with it is that the conjunction over  $H$  is actually infinite. Luckily, there are only finitely many pairwise inequivalent first-order sentences about graphs of quantifier depth  $d$ ; see, e.g., [22, Theorem 2.4]. Removing all but one formula  $\Phi_{G,H}$  from each equivalence class, we make  $\Phi_G$  a legitimate finite sentence. Now, consider  $\Phi \stackrel{\text{def}}{=} \bigvee_G \Phi_G$ , where the disjunction is over all  $G \in \mathcal{C}$ . It can be made finite in the same way. The sentence  $\Phi$  defines  $\mathcal{C}$  and has quantifier depth  $d$ .  $\square$

Thus, Lemma 2.1 is a simple consequence of the fact that there are only finitely many pairwise inequivalent first-order statements of bounded quantifier depth. Note that the last fact does not hold true for the variable width. We define

$$W^*(\mathcal{C}) = \max \{ W(G, H) : G \in \mathcal{C}, H \notin \mathcal{C} \}.$$

Equivalently,  $W^*(\mathcal{C})$  is equal to the minimum  $k$  such that  $\mathcal{C}$  is definable in the infinitary logic  $L_{\infty\omega}^k$ ; see [18]. Obviously,  $W^*(\mathcal{C}) \leq W(\mathcal{C})$ , and we will see in Section 4.2 that this inequality can be strict. Summarizing, we have

$$W^*(\mathcal{C}) \leq W(\mathcal{C}) \leq D(\mathcal{C}). \tag{1}$$

The value of  $W(\mathcal{C})$  admits the following characterization. If  $W(G, H) \leq k$ , let  $D^k(G, H)$  denote the minimum quantifier depth of a first-order  $k$ -variable sentence distinguishing  $G$  and  $H$ . We set  $D^k(G, H) = \infty$  if  $W(G, H) > k$ . The following fact can be proved similarly to Lemma 2.1.

**Lemma 2.2.**  $W(\mathcal{C}) > k$  if and only if there is a sequence of graph pairs  $(G_i, H_i)$  with  $G_i \in \mathcal{C}$  and  $H_i \notin \mathcal{C}$  such that  $D^k(G_i, H_i) \rightarrow \infty$  as  $i \rightarrow \infty$ .

Lemmas 2.1 and 2.2 reduce estimating the logical depth and width to estimating the parameters  $D(G, H)$  and  $D^k(G, H)$  over  $G \in \mathcal{C}$  and  $H \notin \mathcal{C}$ . The first inequality in (1) can be used for obtaining lower bounds for  $W(\mathcal{C})$  by estimating  $W(G, H)$  over  $G \in \mathcal{C}$  and  $H \notin \mathcal{C}$ . The parameters  $D(G, H)$ ,  $D^k(G, H)$ , and  $W(G, H)$  have a very useful combinatorial characterization.

In the  $k$ -pebble *Ehrenfeucht-Fraïssé game*, the board consists of two vertex-disjoint graphs  $G$  and  $H$ . Two players, *Spoiler* and *Duplicator* (or *he* and *she*) have equal sets of  $k$  pairwise different pebbles. In each round, *Spoiler* takes a pebble and puts it on a vertex in  $G$  or in  $H$ ; then *Duplicator* has to put her copy of

this pebble on a vertex of the other graph. Duplicator's objective is to ensure that the pebbling determines a partial isomorphism between  $G$  and  $H$  after each round; when she fails, she immediately loses. The proof of the following facts can be found in [12]:

1.  $D^k(G, H)$  is equal to the minimum  $d$  such that Spoiler has a winning strategy in the  $d$ -round  $k$ -pebble game on  $G$  and  $H$ .
2.  $D(G, H)$  is equal to the minimum  $k$  such that Spoiler has a winning strategy in the  $k$ -round  $k$ -pebble game on  $G$  and  $H$ .
3.  $W(G, H)$  is equal to the minimum  $k$  such that, for some  $d$ , Spoiler has a winning strategy in the  $d$ -round  $k$ -pebble game on  $G$  and  $H$ .

We are interested in the property of containing a specified induced subgraph. We write  $F \sqsubset G$  to say that  $G$  contains an induced subgraph isomorphic to  $F$ . Thus,  $\mathcal{I}(F) = \{G : F \sqsubset G\}$ . Let  $D[F] = D(\mathcal{I}(F))$  and, similarly,  $W[F] = W(\mathcal{I}(F))$  and  $W^*[F] = W^*(\mathcal{I}(F))$ . Thus,  $W^*[F]$  is the maximum  $W(G, H)$  over all  $G$  containing an induced copy of  $F$  and all  $H$  not containing such a copy. As a particular case of (1), we have

$$W^*[F] \leq W[F] \leq D[F] \leq \ell$$

for every  $F$  with  $\ell$  vertices.

The vertex set of a graph  $G$  will be denoted by  $V(G)$ . Let  $\overline{G}$  denote the complement of  $G$ , that is,  $V(\overline{G}) = V(G)$  and two vertices are adjacent in  $\overline{G}$  exactly when they are not adjacent in  $G$ .

**Lemma 2.3.**  $D[F] = D[\overline{F}]$ ,  $W^*[F] = W^*[\overline{F}]$ , and  $W[F] = W[\overline{F}]$ .

*Proof.* The first equality follows from the equality  $D(G, H) = D(\overline{G}, \overline{H})$  by Lemma 2.1. Indeed,  $F \sqsubset G$  iff  $\overline{F} \sqsubset \overline{G}$ . Therefore,

$$\begin{aligned} D[F] &= \max \{ D(G, H) : G \sqsupset F, H \not\sqsupset F \} = \max \{ D(\overline{G}, \overline{H}) : \overline{G} \sqsupset \overline{F}, \overline{H} \not\sqsupset \overline{F} \} \\ &= \max \{ D(G, H) : G \sqsupset \overline{F}, H \not\sqsupset \overline{F} \} = D[\overline{F}]. \end{aligned}$$

The second equality follows similarly from the equality  $W(G, H) = W(\overline{G}, \overline{H})$  by the definition of  $W^*[F]$ . The third equality follows from the equality  $D^k(G, H) = D^k(\overline{G}, \overline{H})$  by Lemma 2.2.  $\square$

**Further graph-theoretic definitions.** A graph  $G$  is called  $F$ -free if  $F \not\sqsubset G$ . The vertex-disjoint union of graphs  $G$  and  $H$  will be denoted by  $G + H$ . Correspondingly,  $sG$  is the vertex-disjoint union of  $s$  copies of  $G$ . The *lexicographic product*  $A \cdot B$  of two graphs  $A$  and  $B$  is defined as follows:  $V(A \cdot B) = V(A) \times V(B)$ , and  $(u, v)$  and  $(x, y)$  are adjacent in  $A \cdot B$  if  $u$  and  $x$  are adjacent in  $A$  or if  $u = x$  and  $v$  and  $y$  are adjacent in  $B$ . In other words,  $A \cdot B$  is obtained from  $A$  by substituting each

vertex  $u$  with an induced copy  $B_u$  of  $B$  and drawing all edges between  $B_u$  and  $B_x$  whenever  $u$  and  $x$  are adjacent.

A vertex is *isolated* if it has no adjacent vertex and *universal* if it is adjacent to all other vertices in the graph. Two vertices are called *twins* if they have the same adjacency to the rest of the graph.

Throughout the paper,  $\log n$  means the logarithm base 2.

### 3 The extension index and a lower bound for $W^*[F]$

Let  $k \geq 2$ . By the  $k$ -extension property we mean the first-order sentence  $EA_k$  of quantifier depth  $k$  (also called the  $k$ -th extension axiom) saying that, for every two disjoint sets  $X, Y \subset V(G)$  with  $|X \cup Y| < k$ , there is a vertex  $z \notin X \cup Y$  adjacent to all  $x \in X$  and non-adjacent to all  $y \in Y$ . Note that  $EA_2$  says exactly that a graph has neither isolated nor universal vertex. For convenience, we also set  $EA_1 \stackrel{\text{def}}{=} \exists z(z = z)$ .

Note that, if  $G \models EA_k$  and  $F$  has at most  $k$  vertices, then  $F \sqsubset G$ . Suppose that  $F$  has more than 1 vertex. We define the *extension index* of  $F$ , denoted by  $E[F]$ , as the minimum  $k$  such that  $H \models EA_k$  implies  $F \sqsubset H$ . Equivalently,  $E[F]$  is the maximum  $k$  for which there is a graph  $H$  such that  $H \models EA_{k-1}$  while  $F \not\sqsubset H$ . Note that  $E[F] \leq \ell$  for any  $\ell$ -vertex graph  $F$ .

**Lemma 3.1.**  $W^*[F] \geq E[F]$ .

*Proof.* As easily seen, if both  $G$  and  $H$  have the  $(k-1)$ -extension property, then Duplicator has a winning strategy in the  $(k-1)$ -pebble Ehrenfeucht-Fraïssé game on graphs  $G$  and  $H$  and, hence,  $W(G, H) \geq k$ . Therefore, it suffices to show that there are  $G$  and  $H$  such that  $F \sqsubset G$ ,  $F \not\sqsubset H$ , and both of them satisfy  $EA_{k-1}$  for  $k = E[F]$ . Such a graph  $H$  exists by the definition of the extension index. Such a graph  $G$  exists because, as very well known (see, e.g., [24]), for fixed  $k$  and  $\ell$  a random graph  $G(n, 1/2)$  has the  $k$ -extension property and contains every  $\ell$ -vertex graph as an induced subgraph with probability approaching 1 as  $n$  increases.  $\square$

**Example 3.2.**

1.  $E[P_3] = 3$  because  $H = 2K_2$  is  $P_3$ -free and satisfies  $EA_2$ . By Lemma 3.1,  $W^*[P_3] = W[P_3] = 3$ .
2.  $E[K_3] = 3$ , as also certified by  $H = 2K_2$  (or by  $H = C_4$ ).

We can determine  $E[K_\ell]$  for any  $\ell$  using a relationship between  $E[F]$  and the chromatic number of  $F$ .

**Theorem 3.3.**  $E[F] \geq \chi(F)$ .

*Proof.* Let  $k = \chi(F) - 1$ . We have to show that there is a graph  $G$  having the  $k$ th extension property and containing no induced copy of  $F$ .

Let  $T_{k,n}$  denote the  $k$ -partite Turán graph with  $kn$  vertices. The vertex set of  $T_{k,n}$  is split into  $k$  vertex classes  $V_1, \dots, V_k$ , each consisting of  $n$  vertices. Two vertices of  $T_{k,n}$  are adjacent if and only if they belong to different vertex classes. Obviously,  $\chi(T_{k,n}) = k$ . Since  $\chi(T_{k,n}) < \chi(F)$ , the graph  $T_{k,n}$  itself and any of its subgraphs do not contain an induced copy of  $F$ . Let  $\mathbb{T}_{k,n}$  be a random subgraph of  $T_{k,n}$ , obtained from  $T_{k,n}$  by deleting each edge with probability  $1/2$ , independently of the other edges. In order to prove the theorem, it suffices to show that  $\mathbb{T}_{k,n}$  has the  $k$ th extension property with nonzero probability if  $n$  is chosen sufficiently large.

Consider two disjoint vertex sets  $X, Y$  in  $\mathbb{T}_{k,n}$  such that  $|X \cup Y| = k - 1$  and estimate the probability that they violate  $EA_k$ . Fix a vertex class  $V_m$  disjoint with  $X \cup Y$ . A particular vertex  $z \in V_m$  is adjacent to all  $x \in X$  and non-adjacent to all  $y \in Y$  with probability  $2^{-k+1}$ , and the converse happens with probability  $1 - 2^{-k+1}$ . The probability that none of the vertices in  $V_m$  has the “right” adjacency pattern to  $X$  and  $Y$  is equal to  $(1 - 2^{-k+1})^n$ . Therefore, two sets  $X, Y$  violating  $EA_k$  exist with probability at most

$$\binom{kn}{k-1} 2^{k-1} (1 - 2^{-k+1})^n \leq \left( \frac{kn e}{k-1} \right)^{k-1} 2^{k-1} e^{-2^{-k+1}n} = \left( \frac{k e}{k-1} \right)^{k-1} 2^{k-1} e^{(k-1) \ln n - 2^{-k+1}n},$$

which approaches 0 as  $n$  increases (since  $k$  is fixed). It follows that  $EA_k$  is violated by  $\mathbb{T}_{k,n}$  with probability strictly less than 1 if  $n$  is chosen sufficiently large.  $\square$

**Corollary 3.4.**  $E[K_\ell] = \ell$ .

It may seem plausible at first glance that  $E[F] = \ell$  for every  $F$  with  $\ell$  vertices. Nevertheless, in Section 4 we will see that this is not always the case as, for example,  $E[F] = 3$  for  $F$  being the paw and the path on 4 vertices. The best general lower bound for  $E[F]$  we can show is given by the following lemma.

**Lemma 3.5.** *Let  $F$  be a graph with  $\ell \geq 2$  vertices. Then*

$$E[F] \geq \lfloor \frac{1}{2} \ell - 2 \log_2 \ell + 3 \rfloor.$$

*Proof.* The lemma is trivially true if  $\ell \leq 15$  because in this case it just states that  $E[F] \geq 2$ . We, therefore, suppose that  $\ell \geq 16$ .

Denote  $k = \lfloor \frac{1}{2} \ell - 2 \log_2 \ell + 2 \rfloor$ . Suppose that  $\ell$  is even and set  $n = 2^{\ell/2-1}$ . It suffices to show that the random graph  $G(n, 1/2)$  with a non-zero probability has the  $k$ -extension property and simultaneously contains no induced copy of  $F$ .

The probability of  $F \sqsubset G(n, 1/2)$  is bounded from above by the value of

$$p(\ell, n) = n(n-1)(n-2) \cdots (n-\ell+1) 2^{-\ell(\ell-1)/2}.$$

Since

$$2^{-\ell(\ell-1)/2} = (2n)^{-\ell+1} = 2^{-\ell+1} n^{-\ell+1} = \frac{1}{2} n^{-\ell-1},$$



we have

$$p(\ell, n) = \frac{n(n-1)(n-2) \cdots (n-\ell+1)}{2n^{\ell+1}} < \frac{1}{2n}.$$

It remains to prove that  $G(n, 1/2)$  has the  $k$ -extension property with probability at least  $1/(2n)$ .

The probability of  $G(n, 1/2) \not\models EA_k$  is bounded from above by the value of

$$q(n, k) = \binom{n}{k-1} 2^{k-1} (1 - 2^{-k+1})^{n-k+1}.$$

In its turn,

$$q(n, k) < 4n^{k-1} (1 - 2^{-k+1})^n.$$

The last value is bounded from above by

$$q'(n, k) = 4 \exp(\ln n(k-1) - n 2^{-k+1}).$$

Denote  $k' = \frac{1}{2}\ell - 2\log \ell + 2$ . Since the function  $f(x) = \ln nx - n 2^{-x}$  is monotonically increasing,

$$\begin{aligned} q(n, k) &< q'(n, k') = 4n^{k'-1} \exp\left(-\frac{n}{2^{\ell/2-2\log \ell+1}}\right) = 4n^{k'-1} \exp\left(-\frac{\ell^2}{4}\right) \\ &= 4n^{k'-1} (2n)^{-\log e \ell/2} = 4n^{k'-1} 2^{-\log e \ell/2} n^{-\log e \ell/2} = 4n^{k'-1} (2n)^{-\log e} n^{-\log e \ell/2} \\ &= 4e^{-1} n^{-\ell(\log e - 1)/2 - 2\log \ell - \log e + 1}. \end{aligned}$$

For  $\ell \geq 16$ , this gives us  $q(n, k) < 2n^{-11}$ . Therefore,  $G(n, 1/2)$  has the  $k$ -extension property with probability more than  $1 - 2n^{-11}$ . This is well more than  $1/(2n)$ , as desired.

If  $\ell$  is odd, set  $n = 2^{(\ell-3)/2}$  and proceed similarly to above.  $\square$

**Remark 3.6.** The bound of Lemma 3.5 cannot be much improved as long as the argument is based on  $G(n, 1/2)$ . Indeed, it is known [14] that there is a function  $\ell_0(n) = 2\log n - 2\log \log n + \Theta(1)$  such that the clique number of  $G(n, 1/2)$  is equal to  $\ell_0(n)$  or to  $\ell_0(n) + 1$  with probability  $1 - o(1)$ . In [15] it is shown that there is a function  $k(n) = \log n - 2\log \log n + \Theta(1)$  such that, with probability  $1 - o(1)$ ,  $G(n, 1/2)$  satisfies  $EA_{k(n)}$  but does not satisfy  $EA_{k(n)+6}$ . It follows that, if  $n$  is chosen so that  $G(n, 1/2)$  does not contain a subgraph  $K_\ell$  with high probability, then  $G(n, 1/2)$  satisfies  $EA_k$  with non-negligible probability for, at best,  $k = \frac{1}{2}\ell - \log \ell + \Theta(1)$ .

As an immediate consequence of Lemmas 3.1 and 3.5 we obtain the following result.

**Theorem 3.7.** *Let  $F$  be a graph with  $\ell \geq 2$  vertices. Then*

$$W^*[F] > \frac{1}{2}\ell - 2\log_2 \ell + 2.$$

## 4 Four-vertex subgraphs

Our next goal is to determine the values of  $D[F]$ ,  $W[F]$ , and  $W^*[F]$  for all graphs  $F$  with at most 4 vertices. It is enough to consider connected  $F$ , as follows from Lemma 2.3 and the fact that the complement of a disconnected graph is connected. The two connected 3-vertex graphs are considered in Example 3.2, and we now focus on connected graphs with 4 vertices. Recall that  $W^*[K_4] = 4$  by Corollary 3.4 (or just because  $W(K_4, K_3) = 4$ ).

### 4.1 The paw subgraph ( $K_3 + e$ )

**Lemma 4.1** (Olariu [20]). *A graph  $H$  is paw-free if and only if each connected component of  $H$  is triangle-free or complete multipartite.*

Note that a graph  $B$  is complete multipartite iff the complement of  $B$  is a vertex-disjoint union of complete graphs. The latter condition means exactly that  $\overline{B}$  is  $P_3$ -free. Thus,  $B$  is complete multipartite iff it is  $(K_2 + K_1)$ -free, where  $K_2 + K_1 = \overline{P_3}$ .

**Theorem 4.2.**  $D[K_3 + e] = W[K_3 + e] = W^*[K_3 + e] = E[K_3 + e] = 3$ .

*Proof.* We have  $E[K_3 + e] > 2$  because, for example,  $C_4$  satisfies the 2nd extension axiom and does not contain  $K_3 + e$  as a subgraph.

In order to prove that  $D[K_3 + e] \leq 3$ , we have to describe a winning strategy for Spoiler in the 3-round Ehrenfeucht-Fraïssé game on graphs  $G \sqsupset K_3 + e$  and  $H \not\supset K_3 + e$ . Let  $v_1, v_2, v_3, v_4$  be vertices spanning a paw in  $G$ . We suppose that  $v_1$  and  $v_2$  have degree 2,  $v_3$  has degree 3, and  $v_4$  has degree 1 in this subgraph. In the first round Spoiler pebbles  $v_1$ . Suppose that Duplicator responds with a vertex  $u_1$  in a connected component  $B$  of  $H$ . By Lemma 4.1,  $B$  is either  $K_3$ -free or a multipartite graph with at least three parts. In the former case Spoiler wins by pebbling  $v_2$  and  $v_3$ . In the latter case Spoiler pebbles  $v_4$  in the second round. The distance between  $v_1$  and  $v_4$  in  $G$  is 2. If Duplicator responds in a connected component of  $H$  other than  $B$ , then he loses in the next round. Therefore, Duplicator is forced in the second round to pebble a vertex  $u_2$  in the same part of  $B$  that contains  $u_1$ . In this case, Spoiler wins by pebbling the vertex  $v_2$ . Indeed, this vertex is adjacent to  $v_1$  and not adjacent to  $v_4$ , while  $u_1$  and  $u_2$  have the same adjacency to any other vertex in  $H$ .  $\square$

### 4.2 The path subgraph ( $P_4$ )

$F = P_4$  is a remarkable example showing that the parameters  $W^*[F]$  and  $W[F]$  can have different values. Specifically, we prove that  $W^*[P_4] = 3$  and  $W[P_4] = 4$ .

**Theorem 4.3.**  $W^*[P_4] = E[P_4] = 3$ .

The proof of Theorem 4.3 is based on a well-known characterization of the class of  $P_4$ -free graphs. A graph is called a *cograph* if it can be built from copies of the

single-vertex graph  $K_1$  by using disjoint unions and complementations. It is known [5] that a graph is  $P_4$ -free if and only if it is a cograph. For the proof of Theorem 4.3 we need the following definitions, that we borrow from [21].

We call  $G$  *complement-connected* if both  $G$  and  $\overline{G}$  are connected. An inclusion-maximal complement-connected induced subgraph of  $G$  will be called a *complement-connected component* of  $G$  or, for brevity, a *cocomponent* of  $G$ . Cocomponents have no common vertices and their vertex sets form a partition of  $V(G)$ . Note that  $G$  is a cograph if and only if all cocomponents of  $G$  are single-vertex graphs.

The *decomposition* of  $G$ , denoted by  $\text{Dec } G$ , is the set of all connected components of  $G$ . Furthermore, given  $i \geq 0$ , we define the *depth  $i$  decomposition*  $\text{Dec}_i G$  of  $G$  by

$$\text{Dec}_0 G = \text{Dec } G \quad \text{and} \quad \text{Dec}_{i+1} G = \bigcup_{E \in \text{Dec}_i G} \text{Dec } \overline{E}.$$

Note that

$$\Pi_i = \{V(E) : E \in \text{Dec}_i G\} \tag{2}$$

is a partition of  $V(G)$ , and  $\Pi_{i+1}$  refines  $\Pi_i$ . Once the partition stabilizes, that is,  $\Pi_{i+1} = \Pi_i$ , it coincides with the partition of  $G$  into its cocomponents. The *depth  $i$  environment* of a vertex  $v \in V(G)$ , denoted by  $\text{Env}_i(v)$ , is the graph  $E$  in  $\text{Dec}_i G$  containing  $v$ .

**Lemma 4.4.** *Suppose that a graph  $G$  contains an induced copy of  $P_4$  and let  $U \subseteq V(G)$  be such that  $G[U] \cong P_4$ . Consider the 3-pebble Ehrenfeucht-Fraïssé game on  $G$  and another graph  $H$ . Let  $x, x' \in V(G)$  and  $y, y' \in V(H)$ , and assume that the pairs  $x, y$  and  $x', y'$  were selected by the players in the same rounds. If  $x, x' \in U$  and  $\text{Env}_l(y) \neq \text{Env}_l(y')$ , then Spoiler has a strategy allowing him to win in this position playing all the time in  $U$  and making no more than  $2l + 2$  moves.*

*Proof.* We use induction on  $l$ . In the base case of  $l = 0$ , the vertices  $y$  and  $y'$  lie in different connected components of  $H$ , while the distance between  $x$  and  $x'$  in  $G$  is at most 3. Therefore, Spoiler is able to win with one extra pebble in 2 moves.

Let  $l \geq 1$ . Suppose that  $\text{Env}_{l-1}(y) = \text{Env}_{l-1}(y') = E$  (for else Spoiler wins by the induction assumption). Note that  $\text{Env}_l(y)$  and  $\text{Env}_l(y')$  are connected components of  $\overline{E}$ . Since  $P_4$  is self-complementary,  $\overline{G[U]} \cong P_4$ . Therefore, if Duplicator moves only in  $V(E)$ , Spoiler will win in at most 2 next moves. Once Duplicator makes one of these moves outside  $V(E)$ , this creates a position with two vertices  $\tilde{x}$  and  $\tilde{x}'$  pebbled by Spoiler in  $U$  such that  $\text{Env}_{l-1}(\tilde{y}) \neq \text{Env}_{l-1}(\tilde{y}')$  for the corresponding vertices  $\tilde{y}$  and  $\tilde{y}'$  pebbled by Duplicator in  $H$ . The induction assumption applies.  $\square$

*Proof of Theorem 4.3.* Consider graphs  $G \sqsupset P_4$  and  $H \not\supset P_4$ . Since  $H$  is a cograph,  $\text{Dec}_i H$  for some  $i$  consists of single-vertex graphs. By Lemma 4.4, this readily implies that  $W(G, H) \leq 3$ . Indeed, when Spoiler pebbles two vertices on an induced  $P_4$  in  $G$ , then whatever Duplicator responds, this creates a position as in Lemma 4.4. Thus,  $W^*[P_4] \leq 3$ .

The lower bound  $E[P_4] > 2$  is certified, for example, by  $C_4$ .  $\square$

Theorem 4.3 implies that the class of graphs containing an induced  $P_4$  is definable in the infinitary logic  $L^3_{\infty\omega}$ . It turns out that this class is not definable in 3-variable first-order logic.

**Theorem 4.5.**  $W[P_4] = 4$ .

Our proof of Theorem 4.5 is based on Lemma 2.2. It suffices to exhibit a sequence of graph pairs  $G_i, H_i$  such that  $G_i$  contains an induced copy of  $P_4$ ,  $H_i$  does not, and  $D^3(G_i, H_i) \rightarrow \infty$  as  $i$  increases.

Given a graph  $A$ , we define its  $i$ -th power  $(A)^i$  by  $(A)^1 = A$  and  $(A)^{i+1} = \overline{(A)^i + (A)^i}$ . Now, let  $H_i = (K_1)^i$ . This is a cograph and, therefore,  $P_4 \not\sqsubset H_i$  (which is also easy to see directly, using induction and the fact that  $P_4$  is self-complementary).

In order to construct  $G_i$ , we use the lexicographic product of graphs; see Section 2. Fix a graph  $A$  satisfying the 3rd extension axiom and containing  $P_4$  as an induced subgraph (a large enough random graph has both of these properties with high probability). Now, let  $G_i = H_i \cdot (A \cdot H_i)$ . Obviously,  $P_4 \sqsubset G_i$ . Theorem 4.5 immediately follows by Lemma 2.2 from the following estimate.

**Lemma 4.6.**  $D^3(G_{4m+2}, H_{4m+2}) > m + 1$ .

The proof of Lemma 4.6 is lengthy and occupies a separate subsection.

### 4.3 Proof of Lemma 4.6

We have to describe a strategy of Duplicator in the 3-pebble game on  $G = G_{4m+2}$  and  $H = H_{4m+2}$  allowing her to win  $m + 1$  rounds. This will be convenient to do in terms of a nonstandard metric on  $H$ , which we introduce now.

Let  $t = 4m + 1$ ; thus,  $H = H_{t+1}$ . Consider the partitions of  $\Pi_0, \dots, \Pi_t$  of  $V(H)$  defined by (2). Note that  $\Pi_0$  is the trivial partition with the single partition element  $V(H)$ , and  $\Pi_t$  is the complete partition of  $V(H)$  into singletons. Let  $x$  and  $y$  be vertices of  $H$ . Define  $\bar{d}(x, y)$  to be the maximum  $k$  such that  $x$  and  $y$  belong to the same element of  $\Pi_k$  or, equivalently,  $\text{Env}_k(x) = \text{Env}_k(y)$ . Furthermore, we set  $d(x, y) = t - \bar{d}(x, y)$ . Note that  $0 \leq d(x, y) \leq t$ . As easily seen,  $d(x, y) = 0$  iff  $x = y$ , and  $d(x, y) = 1$  iff  $x$  and  $y$  are twins.

Note that adjacency between two vertices  $x$  and  $y$  is completely determined by  $d(x, y)$ , and it changes once  $d(x, y)$  increases or decreases by 1.

*Claim A.*

1.  $H$  is vertex-transitive.
2. Every automorphism  $\alpha$  of  $H$  preserves the function  $d$ , i.e.,  $d(\alpha(x), \alpha(y)) = d(x, y)$  for any  $x, y \in V(H)$ .
3. Let  $x, y, x', y' \in V(H)$ . There exists an automorphism  $\alpha$  of  $H$  such that  $\alpha(x) = x'$  and  $\alpha(y) = y'$  if and only if  $d(x, y) = d(x', y')$ .

*Proof of Claim A.* The claim holds true for every  $H = H_i$  and follows by a simple induction on  $i$ . Part 1 follows from an observation that  $H_i = \overline{H_{i-1}} + \overline{H_{i-1}}$  has an automorphism transposing the two connected components of  $\overline{H_i}$ . Part 2 is true by the definition of  $d$  because every automorphism of  $H_j$  preserves connected components of  $\overline{H_j}$  for each  $j$  and, therefore, permutes the elements of each partition  $\Pi_j$ . In order to prove Part 3, consider two cases. First, suppose that  $x$  and  $y$  are in different  $H_{i-1}$ -components of  $\overline{H_i}$ , that is,  $d(x, y) = t$ . If  $d(x, y) = d(x', y')$ , the vertices  $x'$  and  $y'$  are also in different  $H_{i-1}$ -components, and we use an automorphism transposing these components and the induction assumption that  $H_{i-1}$  is vertex-transitive. If  $d(x, y) \neq d(x', y')$ , then such an automorphism does not exist by Part 2. The case that  $x'$  and  $y'$  are in different  $H_{i-1}$ -components of  $\overline{H_i}$  is symmetric. There remain the cases that all four  $x, y, x', y'$  are in the same  $H_{i-1}$ -component of  $\overline{H_i}$  or that  $x, y$  are in one  $H_{i-1}$ -component of  $\overline{H_i}$  and  $x', y'$  are in the other component. They are covered by the induction assumption.  $\triangleleft$

*Claim B.*  $d(x, y) = d(x, z) = d(y, z)$  is impossible for any pairwise distinct  $x, y, z \in V(H)$ .

*Proof of Claim B.* These equalities would imply that  $\bar{d}(x, y) = \bar{d}(x, z) = \bar{d}(y, z) = k$  for some  $k < t$ . This would mean that  $Env_{k+1}(x)$ ,  $Env_{k+1}(y)$ , and  $Env_{k+1}(z)$  are pairwise disjoint while  $Env_k(x) = Env_k(y) = Env_k(z)$ . This is a contradiction because a depth- $k$  environment has only two depth- $(k+1)$  subenvironments.  $\triangleleft$

*Claim C.* If  $d(x, y) < d(x, z)$ , then  $d(y, z) = d(x, z)$ .

*Proof of Claim C.* Denote  $k = \bar{d}(x, y)$ . Then  $Env_k(y) = Env_k(x)$  has an empty intersection with  $Env_k(z)$ . It follows that  $Env_m(z) = Env_m(y)$  iff  $Env_m(z) = Env_m(x)$ .  $\triangleleft$

Claims B and C imply that, with respect to the distance function  $d$ , any three vertices of  $H$  form an isosceles triangle where the base is shorter than the two legs. This implies that  $d$  is a graph metric because

$$d(x, y) \leq \max(d(x, z), d(z, y)) \quad (3)$$

and, hence,  $d(x, y) \leq d(x, z) + d(z, y)$ .

Assume that, in the course of the 3-pebble game on  $G$  and  $H$ , the players pebble vertices  $x$  and  $y$  in  $G$  and the corresponding vertices  $x'$  and  $y'$  in  $H$  so that the position is non-losing for Duplicator. Notice two configurations that are potentially dangerous for Duplicator. If  $d(x', y') = 1$ , then Duplicator loses whenever Spoiler pebbles a vertex  $z$  in  $G$  adjacent to exactly one of  $x$  and  $y$ . If  $d(x', y') = t$ , then Duplicator loses whenever Spoiler pebbles a vertex  $z$  in  $G$  adjacent neither to  $x$  nor to  $y$ . Any other configuration is non-losing for Duplicator if Spoiler moves in  $G$ . In particular, we will use the following fact.

*Claim D.* Suppose that  $2 < d(x', y') < t - 2$ . Then, whatever vertex  $z$  is pebbled by Spoiler in  $G$ , Duplicator has a non-losing move in  $H$ . Moreover, there is a vertex

$z' \in V(H)$  with any desired adjacency to  $x$  and to  $y$  such that

$$|d(z', x') - d(x', y')| \leq 2 \text{ and } |d(z', y') - d(x', y')| \leq 2. \quad (4)$$

*Proof of Claim D.* Denote  $k = \bar{d}(x', y')$ . Assume that  $x' \sim y'$  and, hence  $x \sim y$  (the case that  $x' \not\sim y'$  is symmetric). If  $z \sim x$  and  $z \not\sim y$ , then Duplicator pebbles a vertex  $z'$  in  $\text{Env}_{k+1}(y')$ . This will ensure that  $z' \sim x'$ . More precidely, Duplicator chooses  $z' \in \text{Env}_{k+1}(y') \setminus \text{Env}_{k+2}(y')$  to ensure also that  $z' \not\sim y'$ . Note that  $\bar{d}(z', x') = k$  and  $\bar{d}(z', y') = k + 1$ . This implies (4). The case that  $z \not\sim x$  and  $z \sim y$  is symmetric.

If  $z \not\sim x$  and  $z \not\sim y$ , then Duplicator pebbles a vertex  $z'$  in  $\text{Env}_{k-1}(x') \setminus \text{Env}_k(x')$  (these environments are the same for  $y'$ ). This ensures that  $z' \not\sim x'$  and  $z' \not\sim y'$ . Moreover,  $\bar{d}(z', x') = \bar{d}(z', y') = k - 1$ . If  $z \sim x$  and  $z \sim y$ , then Duplicator pebbles a vertex  $z'$  in  $\text{Env}_{k-2}(x') \setminus \text{Env}_{k-1}(x')$ . Note that  $\bar{d}(z', x') = \bar{d}(z', y') = k - 2$  in this case.  $\triangleleft$

We now make a few useful remarks on the structure of  $G$ .

*Claim E.* For every  $i$ , the graph  $A \cdot H_i$  is complement-connected.

*Proof of Claim E.* By the 3-extension property, both  $A$  and  $\bar{A}$  have diameter 2 and, hence, are connected. Note that  $\overline{A \cdot H_i} = \bar{A} \cdot \bar{H_i}$ . The claim follows from the fact that the lexicographic product is connected whenever the first factor is connected.  $\triangleleft$

Note that  $G_i = H_i \cdot (A \cdot H_i)$  is obtained from  $H_i$  by substituting each vertex for a copy of  $A \cdot H_i$ . It follows by Claim E that all cocomponents of  $G_i$  are isomorphic to  $A \cdot H_i$ . This also shows that, if  $i$  is even (like in our case of  $i = 4m + 2$ ), then  $G_i$  has the same recursive structure as  $H_i$ , namely  $G_i = (A \cdot H_i)^i$ .

Thus, each cocomponent of  $G$  is isomorphic to  $A \cdot H$  and, therefore, consists of  $a$  copies of  $H$  (and edges between them), where  $a$  is the number of vertices in  $A$ . On each of the  $a$  copies of  $H$  we consider the metric  $d$  introduced above. We also consider this metric on the first factor of  $H$  in  $G = H \cdot (A \cdot H)$ , using the different notation  $D$  for it. Moreover, we extend the distance functions  $D$  and  $d$  to the entire set  $V(G)^2$  as follows. Recall that a vertex of  $G$  is a triple  $(h, a, h')$  where  $h, h' \in V(H)$  and  $a \in A$ . Then

$$D((h_1, a_1, h'_1), (h_2, a_2, h'_2)) = d(h_1, h_2)$$

and

$$d((h_1, a_1, h'_1), (h_2, a_2, h'_2)) = \begin{cases} d(h'_1, h'_2) & \text{if } (h_1, a_1) = (h_2, a_2), \\ \infty & \text{otherwise.} \end{cases}$$

Similarly to the dangerous configurations for Duplicator that are mentioned above, note that Duplicator can lose in the next round if two vertices  $x$  and  $y$  are pebbled in  $G$  such that  $d(x, y) = 1$  or  $D(x, y) = t$ . On the other hand, a configuration  $x, y \in V(G)$  with  $d(x, y) = t$  is not dangerous because if Spoiler pebbles  $z' \in V(H)$  with the same adjacency to  $x'$  and  $y'$ , then Duplicator can respond with  $z$  in the same  $A \cdot H$ -cocomponent of  $G$  but in a different copy of  $H$  (not containing  $x$  and

$y$ ). She can ensure adjacency or non-adjacency to both  $x$  and  $y$  because  $A$  has the 3-extension property. A configuration with  $D(x, y) = 1$  also does not pose any threat to Duplicator because she can find a vertex  $z$  adjacent to exactly one of  $x$  and  $y$  either in the  $H$ -part of a cocomponent of  $G$  containing  $x$  or in the  $H$ -part of a cocomponent of  $G$  containing  $y$  (that is, either  $d(z, x) < \infty$  or  $d(z, y) < \infty$ ).

Speaking of a *2-pebble configuration*  $x/x', y/y'$ , where  $x, y \in V(G)$  and  $x', y' \in V(H)$ , we mean that the vertices  $x$  and  $y$  are pebbled in  $G$  and the corresponding pebbles occupy the vertices  $x'$  and  $y'$  in  $H$ . Let  $0 \leq s \leq 2m$ . We call a configuration  $x/x', y/y'$  *s-safe* (for Duplicator) if the following three conditions are true:

- $x \sim y$  iff  $x' \sim y'$ ;
- $d(x, y) = d(x', y')$  or both the distances are sufficiently large:  $d(x, y) > s$  and  $d(x', y') > s$ .
- $D(x, y) = d(x', y')$  or both the distances are sufficiently small:  $D(x, y) < t - s$  and  $d(x', y') < t - s$ .

A 3-pebble configuration  $x/x', y/y', z/z'$  is *s-safe* if every 2-pebble subconfiguration of it is *s-safe*.

We are now ready to describe Duplicator's strategy.

*1st round.* Duplicator's response is arbitrary. If she moves in  $H$ , recall that  $H$  is vertex-transitive. If she moves in  $G$ , there can be non-isomorphic choices, but a particular choice does not influence the game because  $G$  is constructed from the vertex-transitive graph  $H$  and the graph  $A$  having the 3-extension property. Suppose that the players have pebbled vertices  $x \in V(G)$  and  $x' \in V(H)$ .

*2nd round.* The vertices pebbled in this round will be denoted by  $y$  and  $y'$ . Duplicator ensures a  $(2m - 1)$ -safe configuration  $x/x', y/y'$  as follows:

- If  $d(x, y) \leq 2m - 1$  or  $d(x', y') \leq 2m - 1$  (depending on in which of the graphs Spoiler moves), then Duplicator responds so that  $d(x, y) = d(x', y')$ . Note that this guarantees that  $x \sim y$  iff  $x' \sim y'$ .
- If  $D(x, y) \geq 2m + 2$  or  $d(x', y') \geq 2m + 2$  (depending on in which of the graphs Spoiler moves), then Duplicator responds so that  $D(x, y) = d(x', y')$ . Again, this guarantees that  $x \sim y$  iff  $x' \sim y'$ .
- Otherwise, Duplicator takes care that none of the above conditions is true. If moving in  $H$ , she achieves this by pebbling a vertex  $y'$  such that  $d(x', y') = 2m$  or  $d(x', y') = 2m + 1$ , depending on whether or not  $x$  and  $y$  are adjacent. If moving in  $G$ , Duplicator can achieve the desired adjacency between  $x$  and  $y$ , for example, by pebbling  $y$  such that  $d(x, y) = \infty$  and  $D(x, y) = 0$  (due to the 2-extension property of  $A$ ).

The core of Duplicator's strategy in further rounds is stated in the following claim.

*Claim F.* Let  $2 \leq s \leq 2m - 1$ . Suppose that  $x/x', y/y'$  is an  $s$ -safe 2-pebble configuration. Then, whatever Spoiler does in the next round, Duplicator can respond so that the resulting configuration  $x/x', y/y', z/z'$  is  $(s - 2)$ -safe.

Claim F readily implies that Duplicator survives not only in the first and the second rounds but also in at least  $m - 1$  subsequent rounds, which yields the bound in Lemma 4.6. It remains to prove this claim.

*Proof of Claim F.*

*Case A:*  $d(x', y') \leq s$ . Since the configuration  $x/x', y/y'$  is  $s$ -safe, we have  $d(x, y) = d(x', y')$  and, hence,  $x \sim y$  iff  $x' \sim y'$ . We split our analysis into several subcases depending on the mutual position of the vertices  $z$  and  $x$  (or  $z'$  and  $x'$ ). The mutual position of the vertices  $z$  and  $y$  (or  $z'$  and  $y'$ ) will be determined in each of the subcases because the vertices  $x$  and  $y$  (and  $x'$  and  $y'$ ) are close to each other in the metric  $d$ .

*Subcase A-1:* Spoiler pebbles  $z \in V(G)$  such that  $d(z, x) \leq s$ . It follows by (3) that  $d(z, y) \leq s$ . Let  $H'$  be the  $H$ -part of a cocomponent of  $G$  containing the vertices  $x$  and  $y$ . By Claim A, there is an isomorphism  $\alpha$  from  $H'$  to  $H$  such that  $\alpha(x) = x'$  and  $\alpha(y) = y'$ . Duplicator pebbles the vertex  $z' = \alpha(z)$ . The adjacencies are respected automatically. Since

$$\begin{aligned} d(z', x') = d(z, x) &\leq s < t - s \quad \text{and} \\ d(z', y') = d(z, y) &\leq s < t - s, \end{aligned}$$

the configuration  $x/x', y/y', z/z'$  is  $s$ -safe and, hence,  $(s - 2)$ -safe.

*Subcase A-1':* Spoiler pebbles  $z' \in V(H)$  such that  $d(z', x') \leq s$ . Duplicator responds in  $G$  mirroring Subcase A-1.

*Subcase A-2:* Spoiler pebbles  $z \in V(G)$  such that  $D(z, x) \geq t - s$ . Since  $x$  and  $y$  are in the same cocomponent and  $z$  is in a different cocomponent of  $G$ , we have  $D(z, x) = D(z, y)$ , and  $z \sim x$  iff  $z \sim y$ . Duplicator pebbles  $z' \in V(H)$  such that  $d(z', x') = D(z, x)$ . This implies that  $z' \sim x'$  iff  $z \sim x$ . Since  $d(z', x') \geq t - s > s \geq d(x', y')$ , Claim C implies that  $d(z', y') = d(z', x') = D(z, x) = D(z, y)$ . As a consequence,  $z' \sim y'$  iff  $z \sim y$ . It follows that the resulting configuration is  $s$ -safe.

*Subcase A-2':* Spoiler pebbles  $z' \in V(H)$  such that  $d(z', x') \geq t - s$ . Duplicator responds in  $G$  mirroring Subcase A-2.

*Subcase A-3.1:* Spoiler pebbles  $z \in V(G)$  such that  $s < d(z, x) < \infty$ . Since  $d(z, y) < \infty$  and  $d(z, x) > d(x, y)$ , we have  $d(z, y) = d(z, x)$  by Claim C, and  $z \sim y$  iff  $z \sim x$ . Duplicator responds with  $z' \in V(H)$  such that  $d(z', x') = 2m$  or  $d(z', x') = 2m + 1$ . More specifically, she chooses the option ensuring that  $z' \sim x'$  iff  $z \sim x$ . In either case,  $d(z', x') > d(x', y')$  and, hence,  $d(z', y') = d(z', x')$ . Therefore,

$$z' \sim y' \iff z' \sim x' \iff z \sim x \iff z \sim y,$$



and the new configuration remains  $s$ -safe.

*Subcase A-3.2: Spoiler pebbles  $z \in V(G)$  such that  $0 < D(z, x) < t - s$ .* Note that  $D(z, y) = D(z, x)$ . Duplicator plays like in Subcase A-3.1.

*Subcase A-3.3: Spoiler pebbles  $z \in V(G)$  such that  $d(z, x) = \infty$  and  $D(z, x) = 0$ .* Note that  $d(z, y) = \infty$  and  $D(z, y) = 0$ , which implies that  $z \sim y$  iff  $z \sim x$ . Duplicator plays like in Subcase A-3.1.

*Subcase A-3': Spoiler pebbles  $z' \in V(H)$  such that  $s < d(z', x') < t - s$ .* Since  $d(z', x') > s \geq d(x', y')$ , we have  $d(z', y') = d(z', x')$  and, hence,  $z' \sim y'$  iff  $z' \sim x'$ . Duplicator can, for example, mirror Subcase A-3.3 by pebbling  $z \in V(G)$  such that  $d(z, x) = \infty$ ,  $D(z, x) = 0$ , and  $z \sim x$  iff  $z' \sim x'$ . The last equivalence can be achieved due to the 2-extension property of the graph  $A$ . The equivalence  $z \sim y \iff z' \sim y'$  will be true automatically.

*Case B:  $d(x', y') \geq t - s$ .* Since the configuration  $x/x', y/y'$  is  $s$ -safe,  $D(x, y) = d(x', y')$  and  $x \sim y$  iff  $x' \sim y'$  in this case. We split analysis of this case into several subcases depending on the mutual position of the vertices  $z$  and  $x$  (or  $z'$  and  $x'$ ). There are also symmetric subcases depending on the mutual position of  $z$  and  $y$  (or  $z'$  and  $y'$ ), that are omitted below.

*Subcase B-1: Spoiler pebbles  $z \in V(G)$  such that  $d(z, x) \leq s - 2$ .* Note that  $D(z, y) = D(x, y)$ . Duplicator pebbles  $z'$  in  $H$  such that  $d(z', x') = d(z, x)$ . Since  $d(z', x') \leq s - 2 < t - s \leq d(x', y')$ , we have  $d(z', y') = d(x', y') = D(x, y) = D(z, y)$ , which implies the right adjacencies. As easily seen, the new configuration is  $(s - 2)$ -safe.

*Subcase B-1': Spoiler pebbles  $z' \in V(H)$  such that  $d(z', x') \leq s - 2$ .* This case is, in a sense, mirror-symmetric to Subcase B-1. Note that  $d(z', y') = d(x', y')$ . Duplicator pebbles  $z$  in  $G$  such that  $d(z, x) = d(z', x')$ , which implies that  $D(z, y) = D(x, y)$ .

*Subcase B-2: Spoiler pebbles  $z \in V(G)$  such that  $D(z, x) > s - 2$  and  $D(z, y) > s - 2$ .* Duplicator plays according to an isomorphism from the first  $H$ -factor in  $G = H \cdot (A \cdot H)$  to the graph  $H$  taking  $x$  to  $x'$  and  $y$  to  $y'$ , which exists by Claim A.

*Subcase B-2': Spoiler pebbles  $z' \in V(H)$  such that  $d(z', x') > s - 2$  and  $d(z', y') > s - 2$ .* Duplicator mirrors her strategy from Subcase B-2.

*Subcase B-3: Spoiler plays in  $G$  but differently from Subcases B-1 and B-3.* That is, he pebbles  $z \in V(G)$  such that  $0 < D(z, x) \leq s - 2$  or such that  $D(z, x) = 0$  and  $s - 2 < d(z, x) \leq \infty$ . Since  $D(x, y) \geq t - s > s - 2 \geq D(z, x)$ , we have  $D(z, y) = D(y, x)$  and  $z \sim y$  iff  $y \sim x$ . Duplicator pebbles  $z'$  in  $H$  such that  $d(z', x') = 2m$  or  $d(z', x') = 2m + 1$ ; more specifically, she chooses the option ensuring that  $z' \sim x'$  iff  $z \sim x$ . Since  $d(z', x') < t - s \leq d(x', y')$ , we have  $d(x', y') = d(z', y')$  and, therefore,

$$z' \sim y' \iff x' \sim y' \iff x \sim y \iff z \sim y,$$

yielding the right adjacencies. It is straightforward to check that the configuration  $x/x', y/y', z/z'$  is  $(s - 2)$ -safe.

*Case C:*  $s < d(x', y') < t - s$ . Note that in this case we have  $d(x, y) > s$  and  $D(x, y) < t - s$ .

*Subcase C-1:*  $s < d(x, y) < \infty$ .

*Subcase C-1.1:* Spoiler pebbles  $z \in V(G)$  such that  $d(z, x) \leq s - 2$  or  $d(z, y) \leq s - 2$ . Duplicator pebbles  $z' \in V(H)$  with  $d(z', x') = d(z, x)$  or  $d(z', y') = d(z, y)$  respectively.

*Subcase C-1.1':* Spoiler pebbles  $z' \in V(H)$  such that  $d(z', x') \leq s - 2$  or  $d(z', y') \leq s - 2$ . This is the counterpart of Subcase C-1.1.

*Subcase C-1.2:* Spoiler pebbles  $z \in V(G)$  such that  $D(z, x) = D(z, y) \geq t - s + 2$ . Duplicator responds with  $z' \in V(H)$  such that  $d(z', x') = d(z', y') = D(z, x) = D(z, y)$ .

*Subcase C-1.2':* Spoiler pebbles  $z' \in V(H)$  such that  $d(z', x') = d(z', y') \geq t - s + 2$ . This is the counterpart of Subcase C-1.2.

*Subcase C-1.3:* Spoiler pebbles  $z \in V(G)$  such that  $s - 2 < d(z, x) < \infty$  and  $s - 2 < d(z, y) < \infty$  or the pebbles  $z \in V(G)$  outside the  $H$ -part of a cocomponent of  $G$  containing  $x$  and  $y$  such that  $D(z, x) = D(z, y) < t - s + 2$ . Duplicator succeeds by Claim D.

*Subcase C-1.3':* Spoiler pebbles  $z' \in V(H)$  such that  $s - 2 < d(z', x') < t - s + 2$  and  $s - 2 < d(z', y') < t - s + 2$ . If  $z'$  is adjacent to exactly one of  $x'$  and  $y'$ , then Duplicator has a successful move  $z$  in the  $H$ -part of a cocomponent of  $G$  containing  $x$  and  $y$ ; see the proof of Claim D that ensures that  $d(z, x) \geq d(x, y) - 1 > s - 2$  and  $d(z, y) \geq d(x, y) - 1 > s - 2$ . If  $z'$  is adjacent to both or none of  $x'$  and  $y'$ , then Claim D is non-applicable if  $d(x, y) \geq t - 2$ . In this case, Duplicator has a successful choice of  $z$  in the cocomponent of  $G$  containing  $x$  and  $y$  (due to the 2-extension property of the graph  $A$ ).

*Subcase C-2:*  $d(x, y) = \infty$  and  $D(x, y) = 0$ .

*Subcase C-2.1:* Spoiler pebbles  $z \in V(G)$  such that  $d(z, x) \leq s - 2$  or  $d(z, y) \leq s - 2$ . Duplicator pebbles  $z' \in V(H)$  with  $d(z', x') = d(z, x)$  or  $d(z', y') = d(z, y)$  respectively.

*Subcase C-2.1':* Spoiler pebbles  $z' \in V(H)$  such that  $d(z', x') = d(z, x)$  or  $d(z', y') = d(z, y)$ . This is the counterpart of Subcase C-2.1.

*Subcase C-2.2:* Spoiler pebbles  $z \in V(G)$  such that  $D(z, x) = D(z, y) \geq t - s + 2$ . Duplicator responds with  $z' \in V(H)$  such that  $d(z', x') = d(z', y') = D(z, x) = D(z, y)$ .

*Subcase C-2.2':* Spoiler pebbles  $z' \in V(H)$  such that  $d(z', x') = d(z', y') \geq t - s + 2$ . This is the counterpart of Subcase C-2.2.

*Subcase C-2.3:* Spoiler plays in  $G$  but differently from Subcases C-2.1 and C-2.2. Duplicator has an appropriate move in  $H$  by Claim D.

*Subcase C-2.3':* Spoiler pebbles  $z' \in V(H)$  such that  $s - 2 < d(z', x') < t - s + 2$  and  $s - 2 < d(z', y') < t - s + 2$ . Duplicator pebbles a vertex  $z$  in the cocomponent of  $G$  containing  $x$  and  $y$  but in an  $H$ -part of this cocomponent different from the

$H$ -parts containing  $x$  and  $y$ . She is able to obey the adjacency relation due to the 3-extension property of the graph  $A$ .

*Subcase C-3:*  $0 < D(x, y) < t - s$ .

Subcases C-3.1, C-3.1', C-3.2, C-3.2', and C-3.3 are as in Subcase C-2.

*Subcase C-3.3':* Spoiler pebbles  $z' \in V(H)$  such that  $s - 2 < d(z', x') < t - s + 2$  and  $s - 2 < d(z', y') < t - s + 2$ .

If  $d(z', x') = d(z', y') > d(x', y')$ , Duplicator pebbles a vertex  $z$  such that  $D(z, x) = D(z, y)$  is equal either to  $D(x, y) + 1$  or to  $D(x, y) + 2$  so that  $z$  has the required adjacency to  $x$  and  $y$ . Note that  $D(z, x) = D(z, y) < t - s + 2$ , which implies that the configuration is  $(s - 2)$ -safe.

If  $d(z', x') < d(z', y') = d(x', y')$ , Duplicator pebbles a vertex  $z$  such that  $d(z, x) = 2m$  or  $d(z, x) = 2m + 1$ . Making such an “in-depth” move is needed to ensure the  $(s - 2)$ -safety because it is not excluded that  $D(x, y)$  is small, say,  $D(x, y) = 1$ .

The case when  $z'$  is closer to  $y'$  than to  $x'$  in the metric  $d$  is symmetric.  $\triangleleft$

#### 4.4 The claw ( $K_{1,3}$ ) and the diamond ( $K_4 \setminus e$ ) subgraphs

A *strongly regular graph* with parameters  $(n, k, \lambda, \mu)$  is a regular graph with  $n$  vertices of degree  $k$  such that every two adjacent vertices have  $\lambda$  common neighbors and every two non-adjacent vertices have  $\mu$  common neighbors. The simplest examples are  $sK_t$  (the vertex-disjoint union of  $s$  copies of the complete graph  $K_t$ ) and their complements (complete  $s$ -partite graphs with each vertex class of size  $t$ ). We call such strongly regular graphs *trivial*. A strongly regular graph is non-trivial exactly if  $0 < \mu < k < n - 1$ .

An example of a non-trivial strongly regular graph, that will be useful for us below, is the  $m \times m$ -rook graph. The vertex set of this graph is  $\{(a, b) : 1 \leq a, b \leq m\}$ , and two vertices  $(a_1, b_1)$  and  $(a_2, b_2)$  are adjacent if  $a_1 = a_2$  or  $b_1 = b_2$ . In other words, each vertex represents a square of the  $m \times m$  chess board, and two squares are adjacent if one is reachable from the other by a move of the rook. The  $m \times m$ -rook graph is strongly regular with parameters  $(m^2, 2m - 2, m - 2, 2)$ .

The condition  $\lambda = 0$  means that a strongly regular graph is  $K_3$ -free. Every complete bipartite graph  $K_{n,n} = \overline{2K_n}$  has this property and seven other triangle-free non-trivial graphs are known. It is open if there is yet another such graph.

Suppose that  $H$  is a non-trivial non-triangle-free strongly regular graph with parameters  $(n, k, \lambda, \mu)$ . Thus,  $\mu < k$  and it is also not hard to see that  $\lambda < k - 1$  (otherwise every two adjacent vertices were twins and, by connectedness, every two vertices would be adjacent twins, implying that  $H$  is complete). These two inequalities readily imply that  $H$  satisfies the 3rd extension axiom.

**Theorem 4.7.**  $W[K_{1,3}] = W^*[K_{1,3}] = E[K_{1,3}] = 4$  and  $W[K_4 \setminus e] = W^*[K_4 \setminus e] = E[K_4 \setminus e] = 4$ .

*Proof.* By the discussion above, it suffices to exhibit a non-trivial non-triangle-free strongly regular graph that does not contain the claw and the diamond graphs as induced subgraphs. The  $3 \times 3$ -rook graph suits these needs.  $\square$

## 4.5 The cycle subgraph ( $C_4$ )

Let  $G$  be a connected graph. Given  $u, v \in V(G)$ , let  $f_{i,j}^G(u, v)$  denote the number of vertices at distance  $i$  from  $u$  and at distance  $j$  from  $v$ . The graph  $G$  is called *distance-regular* if the number  $f_{i,j}^G(u, v) = f_{i,j}^G(d)$  depends only on  $i, j$ , and the distance  $d = d(u, v)$  between  $u$  and  $v$ . Note that such a graph is regular of degree  $f_{1,1}^G(0)$ . We call two distance-regular graphs  $G$  and  $H$  *similar* if

$$f_{i,j}^G(d) = 0 \iff f_{i,j}^H(d) = 0. \quad (5)$$

**Lemma 4.8.** *If  $G$  and  $H$  are similar distance-regular graphs, then  $W(G, H) > 3$ .*

*Proof.* We show a strategy allowing Duplicator to win the 3-pebble game on  $G$  and  $H$ . In the first round she responds Spoiler's move arbitrarily. Let  $x$  and  $x'$  be the vertices pebbled in  $G$  and  $H$  respectively. Suppose that in the second round Spoiler pebbles a vertex  $y$  in  $G$  (the case that Spoiler plays in  $H$  is similar). Duplicator responds with a vertex  $y'$  in  $H$  such that  $d(x', y') = d(x, y)$ , which guarantees that she does not lose in this round. Such a choice of  $y'$  is possible because (5) implies that  $f_{i,i}^H(0) > 0$  for  $i = d(x, y)$ .

For any subsequent round, assume that  $x, y \in V(G)$  and  $x', y' \in V(H)$  are occupied by two pairs of pebbles and that  $d(x', y') = d(x, y)$ . Suppose that Spoiler puts the third pebble on a vertex  $z$  in  $G$  (the case that Spoiler plays in  $H$  is similar). It is enough to notice that Duplicator can pebble a vertex  $z'$  in  $H$  such that  $d(z', x') = d(z, x)$  and  $d(z', y') = d(z, y)$ . Such a vertex exists due to (5) because  $f_{i,j}^H(d) > 0$  for  $i = d(z, x)$ ,  $j = d(z, y)$ , and  $d = d(x, y)$ .  $\square$

**Theorem 4.9.**  $W[C_4] = W^*[C_4] = 4$ .

*Proof.* By Lemma 4.8, it suffices to exhibit similar distance-regular graphs  $G$  and  $H$  such that  $G$  contains an induced copy of  $C_4$  and  $H$  does not. We can take  $G$  to be the cubic graph (the skeleton of the 3-dimensional cube) and  $H = C_6$ .  $\square$

## 5 Lower bounds over highly connected graphs

As we discussed in Section 1, in the case of a connected pattern graph  $F$  it is algorithmically motivated to consider the parameter  $W_\kappa[F]$ , which is the minimum variable width of a sentence defining the graph class  $\mathcal{I}(F)$  correctly only over graphs of *sufficiently large* connectedness. More precisely,  $W_\kappa[F]$  is equal to the minimum  $k$  for which there is a  $k$ -variable sentence  $\Phi$  and a number  $s$  such that  $G \models \Phi$  iff  $F \sqsubset G$  for all  $s$ -connected graphs  $G$ . Moreover, we define

$$W_\kappa^*[F] = \min_s \max \{ W(G, H) : F \sqsubset G, F \not\sqsubset H, \text{ and both } G \text{ and } H \text{ are } s\text{-connected} \}.$$

This parameter is an analog of  $W_\kappa[F]$  for the infinitary logic, and we have

$$W_\kappa^*[F] \leq W_\kappa[F] \leq W[F] \text{ and } W_\kappa^*[F] \leq W^*[F] \leq W[F].$$

The *join* of graphs  $A$  and  $B$  is denoted by  $A * B$ . Recall that this is the graph obtained from the disjoint union of  $A$  and  $B$  by adding all possible edges between a vertex of  $A$  and a vertex of  $B$ .

**Lemma 5.1.**  $W(A * B, A' * B) \geq W(A, A')$ .

*Proof.* In the game on  $A * B$  and  $A' * B$ , Duplicator can use her strategy for the game on  $A$  and  $A'$ . Each time that Spoiler moves in the  $B$  part of one graph, Duplicator just mirrors his move in the  $B$  part of the other graph.  $\square$

**Lemma 5.2.** Let  $F_0$  be obtained from  $F$  by removing all universal vertices from this graph. Then  $W_\kappa^*[F] \geq W^*[F_0]$ .

*Proof.* Let  $m$  denote the number of universal vertices in  $F$ . Suppose that  $W^*[F_0] = W(G, H)$ , where  $G$  contains an induced copy of  $F_0$  and  $H$  does not. Let  $G'$  be obtained from  $G$  by adding new  $s > m$  universal vertices, and let  $H'$  be defined similarly, i.e,  $G' = G * K_s$  and  $H' = H * K_s$ . Note that  $G'$  contains an induced copy of  $F$  and  $H'$  still does not contain even an induced copy of  $F_0$  (no new vertex of  $H'$  can appear in an induced copy of  $F_0$  because it would be universal there). We have  $W(G', H') \geq W(G, H)$  by Lemma 5.1. This proves the claim because  $G'$  and  $H'$  are  $s$ -connected and  $s$  can be chosen arbitrarily large.  $\square$

**Theorem 5.3.**

1.  $W_\kappa^*[F] = W^*[F]$  whenever  $F$  has no universal vertex.
2.  $W_\kappa^*[F] = W^*[F]$  whenever  $W^*[F] > 3$  and  $F$  has no adjacent twins or no non-adjacent twins.

*Proof.* Part 1 is an immediate consequence of Lemma 5.2. To establish Part 2, we have to prove that  $W_\kappa^*[F] \geq W^*[F]$ . Since with 3 pebbles Spoiler can force playing on connected components, the assumption  $W^*[F] > 3$  implies that  $W^*[F] = W(G, H)$  for some connected  $G$  and  $H$  such that  $F \sqsubset G$  and  $F \not\sqsubset H$ . Assume that  $F$  has no adjacent twins. Consider  $G' = G \cdot K_s$  and  $H' = H \cdot K_s$  (recall that  $\cdot$  denotes the lexicographic product of graphs). Note that  $G'$  and  $H'$  are  $s$ -connected and observe that  $W(G', H') \geq W(G, H)$ . Moreover,  $G'$  still contains an induced copy of  $F$ , and  $H'$  still does not (because if an induced subgraph of  $H'$  contains two vertices from the same  $K_s$ -part, they are adjacent twins in this subgraph). Since  $s$  can be chosen arbitrarily large, this implies that  $W_\kappa^*[F] \geq W(G, H) = W^*[F]$ , as required. If  $F$  has no non-adjacent twins, the same argument works with  $G' = G \cdot \overline{K_s}$  and  $H' = H \cdot \overline{K_s}$ .  $\square$

An example of the graph to which Theorem 5.3 is non-applicable is the diamond. It has universal vertices and both adjacent and non-adjacent twins.

If an  $\ell$ -vertex pattern graph  $F$  has no universal vertex, then  $W_\kappa^*[F] = W^*[F]$  and, therefore,  $W_\kappa^*[F] \geq (\frac{1}{2} - o(1))\ell$  by Theorem 3.7. Lemma 5.2 works well also for  $F$

with few universal vertices. For example, we have  $W_\kappa^*[K_{1,\ell}] \geq W^*[\overline{K_\ell}] = W^*[K_\ell] = \ell$ . However, if  $F$  has many universal vertices, then we need a different approach.

Similarly to  $E[F]$ , we define  $E_\kappa[F]$  to be the maximum  $k$  such that, for each  $s$ , there is an  $s$ -connected graph  $H$  with  $H \models EA_{k-1}$  while  $F \not\subseteq H$ .

The following relations are easy to prove similarly to Lemma 3.1 and Theorem 3.3 (note that, for each fixed  $s$ , the random multipartite graph  $\mathbb{T}_{k,n}$  in the proof of Theorem 3.3 is  $s$ -connected with high probability).

**Lemma 5.4.**  $W_\kappa^*[F] \geq E_\kappa[F] \geq \chi(F)$ .

**Theorem 5.5.** *If  $F$  has  $\ell$  vertices, then  $W_\kappa^*[F] > \frac{1}{3}\ell - \frac{4}{3}\log_2 \ell$ .*

*Proof.* Denote the number of universal vertices in  $F$  by  $m$ , and let  $F_0$  be obtained from  $F$  by removing all these vertices. By Lemma 5.2 and Theorem 3.7,

$$W_\kappa^*[F] \geq W_\kappa^*[F_0] > \frac{1}{2}(\ell - m) - 2\log \ell. \quad (6)$$

By Lemma 5.4,

$$W_\kappa^*[F] \geq \chi(F) \geq m. \quad (7)$$

Combining the bounds (6) and (7), we obtain

$$3W_\kappa^*[F] > \ell - 4\log \ell,$$

which implies the bound stated in the theorem.  $\square$

**Remark 5.6.** In contrast to Lemma 2.3, we cannot deduce just from the definitions that  $W_\kappa^*[F] = W_\kappa^*[\overline{F}]$  because the complement  $\overline{G}$  of a highly connected graph  $G$  can have low connectivity. However, the complement of the random multipartite graph  $\mathbb{T}_{k,n}$  is as well highly connected. In view of this fact, a slight modification of the proof of Theorem 3.3 and Lemma 5.4 gives us also the bound  $W_\kappa^*[F] \geq \chi(\overline{F})$ , which implies, for example, that  $W_\kappa^*[P_\ell] \geq (\ell - 1)/2$  and  $W_\kappa^*[C_\ell] \geq (\ell - 1)/2$ .

Finally, we determine the values of  $W_\kappa^*[F]$  and  $W_\kappa[F]$  for small connected pattern graphs. As a particular case of Lemma 5.4, we have  $W_\kappa^*[K_3] = 3$  and  $W_\kappa^*[K_4] = 4$ . According to the discussion in Section 1, we have  $W_\kappa[P_3] \leq D_v[P_3] \leq 2$ . On the other hand,  $W_\kappa^*[P_3] \geq 2$  just because there are highly connected graphs with an induced copy of  $P_3$  and without it, for example,  $K_n \setminus e$  and  $K_n$  respectively.

**Theorem 5.7.**

1.  $W_\kappa^*[K_3 + e] = W[K_3 + e] = 3$ ;
2.  $W_\kappa^*[P_4] = W^*[P_4] = 3$  and  $W_\kappa[P_4] = W[P_4] = 4$ ;
3.  $W_\kappa^*[F] = W[F] = 4$  for all remaining connected  $F$  on 4 vertices.

*Proof.* 1. In the trivial direction,  $W_\kappa^*[K_3 + e] \leq W[K_3 + e] = 3$ . On the other hand, we have

$$W_\kappa^*[K_3 + e] \geq W^*[K_2 + K_1] \geq E[K_2 + K_1] = E[P_3] = 3,$$

the first inequality due to Lemma 5.2.

2. We have  $W_\kappa^*[P_4] = W^*[P_4] = 3$  by Lemma 5.2. This lemma actually extends to the relation  $W_\kappa[F] \geq W[F_0]$ , and we also have  $W_\kappa[P_4] = W[P_4] = 4$ .

3. We have  $W_\kappa^*[C_4] = W^*[C_4] = 4$  again by Lemma 5.2.

We also have  $W_\kappa^*[K_{1,3}] = E_\kappa[K_{1,3}] = 4$  and  $W_\kappa^*[K_4 \setminus e] = E_\kappa[K_4 \setminus e] = 4$  by the first inequality in Lemma 5.4. Indeed,  $E_\kappa[K_{1,3}] = E_\kappa[K_4 \setminus e] = 4$  because the  $m \times m$  rook's graph is a strongly regular graph containing no induced copy of  $K_{1,3}$  and no induced copy of  $K_4 \setminus e$ . Moreover, the  $m \times m$  rook's graph is  $(2m - 2)$ -connected by the following general fact: The connectivity of a connected strongly regular graph equals its vertex degree [2].  $\square$

Note that the equality  $D_\kappa[K_4 \setminus e] = 4$  follows also from the general lower bound  $D_\kappa[F] \geq \frac{e(F)}{v(F)} + 2$  proved in [25, Theorem 5.4].

## 6 Conclusion

- The equality  $D[K_3 + e] = 3$  is currently the only example we know that shows that  $D[F]$  and  $W[F]$  can be strictly less than the number of vertices in  $F$ . Are there other such graphs? Are there infinitely many of them? On the other hand, we only know that  $W[F] \geq (\frac{1}{2} - o(1))\ell$  for all  $F$  with  $\ell$  vertices. This does not even exclude the possibility that the time bound  $O(n^{W[F]})$  for Induced Graph Isomorphism can be better than the Nešetřil-Poljak bound  $O(n^{(\omega/3)\ell+2})$  for infinitely many pattern graphs  $F$ .
- An example of  $F = P_4$  shows that  $W^*[F]$  can be strictly less than  $W[F]$  but we do not know how far apart from each other  $W^*[F]$  and  $W[F]$  can generally be.
- Note that Lemma 3.1 and Theorem 3.7 show the following hierarchy of graph parameters:

$$(1/2 - o(1))\ell \leq E[F] \leq W^*[F] \leq W[F] \leq D[F] \leq \ell, \quad (8)$$

where  $\ell$  is the number of vertices in  $F$ . It seems that we currently have no example separating the parameters  $W[F]$  and  $D[F]$  or the parameters  $E[F]$  and  $W^*[F]$ . An important question is whether or not  $E[F] = (1 - o(1))\ell$ .

- It follows from (8) that  $W[F] \leq (2 + o(1))W^*[F]$ . In other terms, in the context of Induced Subgraph Isomorphism, the infinitary logic cannot be much more succinct than the standard first-order logic with respect to the number of variables. More generally, is it true that  $W(\mathcal{C}) = O(W^*(\mathcal{C}))$  for all first-order definable graph properties?

- We have checked that  $D[F] = W[F] = \ell$  for all  $F$  with  $\ell \leq 4$  vertices excepting for the paw graph and its complement. Since it remains open if the equality holds true for all larger graphs, it seems reasonable to examine the pattern graphs on 5 vertices. If there is a 5-vertex  $F$  with  $W[F] = 4$ , the resulting decision procedure for  $\text{ISI}(F)$  would be competitive to (or, at least comparable with) the currently known algorithmic results for 5-vertex induced subgraphs. In a more precise form than cited in Section 1, the Nešetřil-Poljak algorithm [19] solves  $\text{ISI}(F)$  in time  $O(n^{\omega a + b})$  if  $F$  has  $3a + b$  vertices. For the graphs on 5 vertices this gives us the time bound  $O(n^{4.373})$ . For some of these 34 graphs, this can be improved to  $O(n^4)$  [9] and an even better bound  $O(n^{3.373})$  can be achieved in some cases [26].
- We have a (rather trivial) example of  $F = P_3$  showing that  $W_\kappa[F]$  can be strictly smaller than  $W[F]$ . Are there other such graphs?

## References

- [1] Noga Alon, Raphael Yuster, and Uri Zwick. Color-coding. *J. ACM*, 42(4):844–856, 1995.
- [2] A.E. Brouwer and D.M. Mesner. The connectivity of strongly regular graphs. *European Journal of Combinatorics*, 6(3):215 – 216, 1985.
- [3] Jianer Chen, Xiuzhen Huang, Iyad A. Kanj, and Ge Xia. Strong computational lower bounds via parameterized complexity. *J. Comput. Syst. Sci.*, 72(8):1346–1367, 2006.
- [4] Yijia Chen and Jörg Flum. On parameterized path and chordless path problems. In *22nd Annual IEEE Conference on Computational Complexity (CCC 2007), 13-16 June 2007, San Diego, California, USA*, pages 250–263. IEEE Computer Society, 2007.
- [5] D. G. Corneil, H. Lerchs, and L. Stewart Burlingham. Complement reducible graphs. *Discrete Appl. Math.*, 3(3):163–174, 1981.
- [6] Derek G. Corneil, Yehoshua Perl, and Lorna K. Stewart. A linear recognition algorithm for cographs. *SIAM J. Comput.*, 14(4):926–934, 1985.
- [7] Bruno Courcelle. The monadic second-order logic of graphs I. Recognizable sets of finite graphs. *Inf. Comput.*, 85(1):12–75, 1990.
- [8] Friedrich Eisenbrand and Fabrizio Grandoni. On the complexity of fixed parameter clique and dominating set. *Theor. Comput. Sci.*, 326(1-3):57–67, 2004.
- [9] Peter Floderus, Mirosław Kowaluk, Andrzej Lingas, and Eva-Marta Lundell. Detecting and counting small pattern graphs. In *Algorithms and Computation*



- *24th International Symposium, ISAAC 2013, Proceedings*, volume 8283 of *Lecture Notes in Computer Science*, pages 547–557. Springer, 2013.
- [10] Peter Floderus, Mirosław Kowaluk, Andrzej Lingas, and Eva-Marta Lundell. Induced subgraph isomorphism: Are some patterns substantially easier than others? *Theoretical Computer Science*, 605:119–128, 2015.
  - [11] François Le Gall. Powers of tensors and fast matrix multiplication. In *International Symposium on Symbolic and Algebraic Computation, ISSAC 2014*, pages 296–303. ACM, 2014.
  - [12] Neil Immerman. *Descriptive complexity*. New York, NY: Springer, 1999.
  - [13] Alon Itai and Michael Rodeh. Finding a minimum circuit in a graph. *SIAM J. Comput.*, 7(4):413–423, 1978.
  - [14] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random graphs*, New York, Berlin: Wiley, 2000.
  - [15] Jeong Han Kim, Oleg Pikhurko, Joel Spencer, and Oleg Verbitsky. How complex are random graphs in first order logic? *Random Struct. Algorithms*, 26(1-2):119–145, 2005.
  - [16] Ton Kloks, Dieter Kratsch, and Haiko Müller. Finding and counting small induced subgraphs efficiently. *Inf. Process. Lett.* 74(3-4):115–121, 2000.
  - [17] Yuan Li, Alexander A. Razborov, and Benjamin Rossman. On the  $AC^0$  complexity of Subgraph Isomorphism. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014*, pages 344–353. IEEE Computer Society, 2014.
  - [18] Leonid Libkin. *Elements of Finite Model Theory*. Texts in Theoretical Computer Science. An EATCS Series. Springer, 2004.
  - [19] Jaroslav Nešetřil and Svatopluk Poljak. On the complexity of the subgraph problem. *Commentat. Math. Univ. Carol.*, 26:415–419, 1985.
  - [20] Stephan Olariu. Paw-free graphs. *Inf. Process. Lett.*, 28:53–54, 1988.
  - [21] Oleg Pikhurko, Joel Spencer, and Oleg Verbitsky. Decomposable graphs and definitions with no quantifier alternation. *Eur. J. Comb.*, 28(8):2264–2283, 2007.
  - [22] Oleg Pikhurko and Oleg Verbitsky. Logical complexity of graphs: a survey. In Martin Grohe and Janos Makowsky, editors, *Model theoretic methods in finite combinatorics*, volume 558 of *Contemporary Mathematics*, pages 129–179. American Mathematical Society (AMS), Providence, RI, 2011.

- [23] Benjamin Rossman. On the constant-depth complexity of  $k$ -clique. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing 2008*, pages 721–730. ACM, 2008.
- [24] J.H. Spencer. *The strange logic of random graphs*. Berlin: Springer, 2001.
- [25] Oleg Verbitsky and Maksim Zhukovskii. The descriptive complexity of subgraph isomorphism without numerics. In *Proc. of the 12th Int. Computer Science Symposium in Russia*, Springer Lecture Notes in Computer Science, 2017. To appear, available as an e-print <http://arxiv.org/abs/1607.08067>.
- [26] Virginia Vassilevska Williams, Joshua R. Wang, Richard Ryan Williams, and Huacheng Yu. Finding four-node subgraphs in triangle time. In *Proceedings of the 26 Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015*, pages 1671–1680. SIAM, 2015.

## A An example: A not necessarily induced $P_4$ subgraph

**Theorem A.1.**  $W_v(P_4) = 2$ , i.e., the existence of a (not necessarily induced) subgraph  $P_4$  is definable over sufficiently large connected graphs by a 2-variable first-order sentence.

*Proof.* Suppose that  $G$  contains a (not necessarily induced) subgraph  $P_4$  and  $H$  does not. Let  $H$  have more than 3 vertices; then  $H = K_{1,n}$ . We have to show that Spoiler wins the 2-pebble game on  $G$  and  $H$  making a bounded number of moves (that does not depend on how large  $G$  and  $H$  are). In the first round he pebbles the central vertex in  $H$ . Suppose that  $G$  has a universal vertex for else Spoiler wins in the next round. Not to lose in the next round, Duplicator pebbles a universal vertex  $u$  in  $G$ . If  $G$  has yet another universal vertex, Spoiler pebbles it and wins in the 3rd round by reusing the first pebble. Assume, therefore, that  $u$  is the only universal vertex in  $G$ . Since  $G$  is not a star graph, it contains a vertex  $v \neq u$  having a neighbor  $w \neq u$ . Spoiler pebbles  $v$  in the second round. Duplicator responds with leaf in  $H$ . In the third round Spoiler moves the pebble from  $u$  to  $w$ . Duplicator is forced to respond with pebbling the central vertex in  $H$ . Spoiler wins in the fourth round by moving the pebble from  $v$  to a vertex non-adjacent with  $w$ .  $\square$